OSU Working Papers in Linguistics 52, 137-149

RULING OUT CHANCE, UNIVERSALITY, AND BORROWING: AN ALTERNATIVE TO RINGE*

GWANG-YOON GOH

1. RINGE'S (1992) PROBABILISTIC METHOD

Ringe (1992) proposes a mathematical method of determining whether the similarities between the basic vocabularies of putatively related languages are the result of chance or not.[1] Although he claims that his method provides "a completely objective criterion of proof" of putative relationships (p.80), his approach has many problems from both linguistic and methodological standpoints some of which are so serious that they render his method partially or even totally incapable of producing accurate or meaningful results.

Above all, although his method can calculate the probability of there being a particular number of matchings of the same kind between the two relevant sounds in a comparable position, he doesn't provide a way of determining how likely a particular number of recurrent matchings (RMs) are to occur by chance. Thus, his method can only give us some sort of strong impression about language relationships. For example, 16 RMs occurring in the first positions of 70 word pairs of an English and German 100-word list might be impressive enough to make anybody believe the close relationship between English and German. However, it cannot answer what this high number (of RMs) means probabilistically or how different numbers of RMs from different pairs of languages can be compared.[2] This problem forces Ringe to appeal to historical arguments to explain the unexpected two RMs found between English and Turkish, making his argument rather circular (pp. 49-50).

* An earlier version of this paper was presented at the 24th annual meeting of Berkeley Linguistics Society, University of California, Berkeley, February 1998. I am grateful to the participants for helpful comments. I would also like to thank Mary Beckman, Steve Keiser, and Craig Hilts for their comments on earlier versions of this paper.

[1] He applies his method to four different pairs of languages, i.e., English-German, English-Latin, English-Turkish, and English-Navajo (Ringe 1992). He also uses his method with some revisions to test the controversial Nostratic hypothesis (Ringe 1995), to debunk Greenberg's multilateral comparison of 'Amerind' family (Ringe 1996), and to suggest a genetic relationship between Indo-European and Uralic language families (1998).

[2] For a more detailed discussion of this problem, see Baxter and Ramer (1996).

Furthermore, Ringe's method can easily result in an undesirable conclusion, because his evidence is based on the number of RMs, whose probabilities cannot be nicely combined. For example, if closely-related languages show a smaller number of RMs than distantly-related languages, his method will give us a wrong prediction about the given relationships. Such a case is actually found in Welby & Whitman (1999 in this volume), which adopts Ringe's method and applies it to three pairs of Algonquian languages: remotely-related Ojibwa and Yurok show 8 RMs, while closely- related Ojibwa and Cree on the one hand and Ojibwa and Arapaho on the other show 3 RMs and 4 RMs, respectively. Such problems can be solved by fully appreciating the wisdom of the traditional comparative method.

In short, Ringe's approach, despite its pioneering role, fails to attain its main goal of computing the chance probability for the relationship between putatively related languages. In this paper, I will propose an alternative method, which bases its probabilistic evidence mainly on the convergence of 'similarities'.

2. MAKING THE BEST OF PROBABILISTIC EVIDENCE

2.1. THE DESCRIPTION OF THE ALTERNATIVE METHOD

For a detailed description and explanation of the general probabilistic methods and procedures involved, which are summarized in (1), I refer to Ringe (1992).

(1) General probabilistic methods and procedures involved

(a) Compile a (Swadesh) word list for the two languages to be compared.
(b) Choose positions for comparison.
(c) Calculate the probability of all possible segment correspondences.
(d) Tabulate the actual matchings.
(e) Calculate the binomial distribution for n, a given number of trials, and p, the probability of a segment correspondence on any trial.[3]
(f) Sort out RMs.

In this paper, "a set of singular facts" and "the convergence of singular facts" in the traditional comparative method (Meillet 1967: 14) are interpreted as a set of RMs and occurrences of multiple matchings [MMs] (i.e. the occurrence of more than one instance of RMs in a given word pair), respectively. Furthermore, the concept 'similarity' is defined probabilistically. Two sounds in a comparable position are 'similar' if their matching turns out to be an RM (i.e. if it falls in the 99th percentile of their expected range), because their correspondence is very difficult to explain unless they are assumed to be reflexes of the same sound.

The alternative method differs from Ringe's mainly in that RMs and their non-chance probabilities, which, in Ringe's method, are intended to be the main evidence for non-chance relationships, are mainly used here to identify 'similar' sounds between two compared languages. Thus, the non-chance probability of RMs, which cannot be

[3] Probability (= P') that there will be **k** matches in **n** random trials, for any number **k**:

• $P' = [n! / k! \times (n - k)!] \times P^k \times (1 - P)^{n-k}$ (cf. Paulos 1988)

n = the total number of trials (i.e. the number of word pairs in a given list).
P = probability of a segment correspondence (i.e. probability that two sounds will match) in a comparable position on any trial.

determined against the given total word list, will be reflected in the calculation of the probability of MMs.

The further steps in the alternative method are as follows: first, using the method employed in Ringe (1992), determine what sounds are 'similar' on the basis of RMs actually found; second, determine how many MMs a given pair of languages show; third, calculate the probability of the convergence of 'similarities', on the basis of the frequencies of the sounds in each comparable position, the 'similar sounds' between two languages, and the number of MMs; finally, provide a probabilistic interpretation of the putative relationships.

2.3. THE PROBABILITY OF THE CONVERGENCE OF 'SIMILARITIES'

The probability that a MM occurs in a word pair can be calculated, as in (2). Moreover, the probability that a particular number of MMs will occur in the given 100-word list can be calculated by the formula for binomial distribution in (3).

(2) Probability [= P(n)] of an n-tuple RM in a word pair

(a) **P(0)** = probability that no RM occurs in any position

$= P\underline{\underline{12345}} = (1-P1) \times (1-P2) \times (1-P3) \times (1-P4) \times (1-P5)$

$= P\underline{\underline{1}} \times P\underline{\underline{2}} \times P\underline{\underline{3}} \times P\underline{\underline{4}} \times P\underline{\underline{5}}$

• **Pn** (or **P**$\underline{\underline{\mathbf{n}}}$) = probability that an RM (or no RM) occurs in **n**-th position

e.g. if the first positions show three RMs: d-t, s-s, k-k,

P1 = (#d/100 x #t/100) + (#s/100 x #s/100) + (#k/100 x #k/100)

(b) **P(1)** = probability that any RM occurs in a word pair.

$= P(2) + (P1\underline{\underline{2345}} + P\underline{\underline{1}}2\underline{\underline{345}} + P\underline{\underline{12}}3\underline{\underline{45}} + P\underline{\underline{123}}4\underline{\underline{5}} + P\underline{\underline{1234}}5) = \mathbf{1 - P(0)}$

• $\mathbf{P1\underline{\underline{2345}}}$ = probability that we have an RM only in the first position.

$= P1 \times (1-P2) \times (1-P3) \times (1-P4) \times (1-P5)$

(c) **P(2)** = probability that any multiple RM occurs in a word pair.

$= P(3) + (P12\underline{\underline{345}} + P1\underline{\underline{2}}3\underline{\underline{45}} + P\underline{\underline{1}}23\underline{\underline{45}} + P\underline{\underline{123}}45 + P\underline{\underline{12}}3\underline{\underline{4}}5 +$
$P\underline{\underline{12}}34\underline{\underline{5}} + P1\underline{\underline{234}}5 + P\underline{\underline{1}}2\underline{\underline{34}}5 + P1\underline{\underline{23}}4\underline{\underline{5}} + P\underline{\underline{1}}2\underline{\underline{3}}4\underline{\underline{5}})$

$= \{1 - P(0) - P''\}$, where **P''** is the probability of any non-multiple RM.

$= (1- P\underline{\underline{12345}} - [P1\underline{\underline{2345}} + P\underline{\underline{1}}2\underline{\underline{345}} + P\underline{\underline{12}}3\underline{\underline{45}} + P\underline{\underline{123}}4\underline{\underline{5}} + P\underline{\underline{1234}}5])$

(d) **P(3)** = probability that any more-than-double RM occurs in a word pair.

$= P(4) + (P\underline{\underline{12}}345 + P\underline{\underline{1}}2\underline{\underline{3}}45 + P\underline{\underline{1}}23\underline{\underline{4}}5 + P\underline{\underline{1}}234\underline{\underline{5}} + P1\underline{\underline{23}}45 +$
$P1\underline{\underline{2}}3\underline{\underline{4}}5 + P1\underline{\underline{2}}34\underline{\underline{5}} + P12\underline{\underline{34}}5 + P12\underline{\underline{3}}4\underline{\underline{5}} + P123\underline{\underline{45}})$

(e) **P(4)** = probability that any more-than-triple RM occurs in a word pair.

$= P(5) + (P\underline{\underline{1}}2345 + P1\underline{\underline{2}}345 + P12\underline{\underline{3}}45 + P123\underline{\underline{4}}5 + P1234\underline{\underline{5}})$

(f) **P(5)** = probability that any quintuple RM occurs in a word pair.

$= P12345 = P1 \times P2 \times P3 \times P4 \times P5$

(3) Probability that a particular number of MMs occur in the given 100-word list

$$\frac{100!}{x!\,(100-x)!} \times (P)^{x} \times (1-P)^{100-x}$$, where **P** = P(n) and **x** = the number of MMs.

3. INVESTIGATION AND RE-INTERPRETATION OF THE DATA

3.1. ENGLISH-GERMAN

The first 100-word list of English and German (cf. Ringe 1992: 83-85) shows many 'similar' sounds in each of the comparable positions, and the corresponding probabilities P1, P2, P3, P4, and P5 (i.e. probability that any real match occurs in each chosen position in a word pair) can be given, as in (4)-(8):[4]

(4) First position [cf. Ringe 1992: 19, 24]

similar sounds	(frequencies)	number of matches
[1] s - s	(14 - 8)	5
[2] s - z	(14 - 7)	6
[3] b - b	(10 - 8)	5
[4] h - h	(9 - 9)	6
[5] Ø - Ø	(8 - 9)	8
[6] n - n	(8 - 5)	5
[7] f - f	(8 - 11)	8
[8] w - v	(7 - 8)	4
[9] l - l	(5 - 5)	4
[10] m - m	(5 - 4)	3
[11] t - c	(5 - 3)	3
[12] k - k	(5 - 7)	3
[13] r - r	(4 - 5)	3
[14] d - t	(4 - 3)	2
[15] g - g	(3 - 5)	3
[16] ð - d	(2 - 2)	2

Total: 16 pairs of similar sounds (= recurrent matchings) in 70 word-pairs

• **P1** = (#s/100 x #s/100) + ... + (#ð/100 x #d/100) = 768/10000 = 0.0768

(5) Second position [cf. Ringe 1992: 25-6]

similar sounds	(frequencies)	number of matches
[1] l - l	(7 - 7)	5
[2] r - r	(5 - 7)	4
[3] t - t	(3 - 4)	3

Total: 3 pairs of similar sounds in 12 word-pairs

• **P2** = (7/100 x 7/100) + (5/100 x 7/100) + (3/100 x 4/100) = 96/10000 = 0.0096

[4] The comparable compositions are determined, as in Ringe (1992): first position = initial consonants; second position = second consonants of the initial clusters; third position = consonants right after the first-syllable vowel nucleus; fourth position = second consonants after the first-syllable vowel nucleus; fifth position = final syllables.

(6) Third position [cf. Ringe 1992: 27-31]

similar sounds	(frequencies)	number of matches
[1] Ø - Ø	(18 - 10)	7
[2] n - n	(15 - 17)	9
[3] t – s	(13 - 11)	8
[4] r - r	(10 - 13)	7
[5] l - l	(8 - 7)	4
[6] d - t	(6 - 7)	3
[7] m - m	(5 - 5)	3
[8] s - z	(3 - 3)	2
[9] s - s	(3 - 3)	3
[10] ŋ - ŋ	(3 - 3)	3
[11] v - b	(2 - 2)	2

Total: 11 pairs of similar sounds in 51 word-pairs

• **P3** = (#Ø/100 x #Ø/100) + ... + (#v/100 x #b/100) = 862/10000 = 0.0862

(7) Fourth position [cf. Ringe 1992: 32-3]

similar sounds	(frequencies)	number of matches
[1] d - d	(6 - 8)	3

Total: 1 pair of similar sounds in 3 word-pairs [5]

• **P4** = 48/10000 = 0.0048

(8) Fifth position (= final syllable) [cf. Ringe 1992: 34-5]

similar sounds	(frequencies)	number of matches
[1] ər - ər	(4 - 4)	4

Total: 1 pair of similar sounds in 4 word-pairs

• **P5** = 16/10000 = 0.0016

On the basis of the probabilities for any RMs in each of the positions determined above, we can calculate the probability for convergence of similarities. For English and German, all the probabilities for all the different kinds of MMs are provided in order to show how each probability is calculated. However, P(2) (= probability that any multiple match occurs in a word pair) alone will often be enough to provide the probabilistic interpretation of the putative relationship. Thus, P(2), P(3), P(4), and P(5) are given in (9), (10), (11), (12), respectively.

(9) P(2) (= probability that any multiple RM occurs in a word pair)

P(2) = P(3) + Probability of any double RM in any positions
= 1 - {P(0) + P"}, where **P"** is the probability of any non-multiple RM.
= $1 - \{P\underline{\underline{12345}} + (P1\underline{\underline{2345}} + P\underline{\underline{1}}2\underline{\underline{345}} + P\underline{\underline{12}}3\underline{\underline{45}} + P\underline{\underline{123}}4\underline{\underline{5}} + P\underline{\underline{1234}}5)\}$
= 1 - 0.990935836 = 0.009064164 < 0.009 < 0.01

[5] This number of matches falls just below the 99th percentile. See Ringe (1992: 33).

$P\underline{\underline{12345}}$ = 0.8301804843; $P1\underline{\underline{2345}}$ = 0.0690618082; $P\underline{\underline{1}}2\underline{\underline{345}}$ = 0.0080469837
$P\underline{\underline{12}}3\underline{\underline{45}}$ = 0.0783120573; $P\underline{\underline{123}}4\underline{\underline{5}}$ = 0.0040040859; $P\underline{\underline{1234}}5$ = 0.0013304174

(10) P(3) (= probability that any more-than-double RM occurs in a word pair)

P(3) = P(4) + Probability of any triple match in a word pair
= P(4) + ($P\underline{\underline{12}}345$ + $P\underline{\underline{1}}2\underline{\underline{3}}45$ + $P\underline{\underline{1}}23\underline{\underline{4}}5$ + $P\underline{\underline{1}}234\underline{\underline{5}}$ + $P1\underline{\underline{23}}45$ +
$P\overline{1}\underline{\underline{2}}3\underline{\underline{4}}5$ + $P\overline{1}\underline{\underline{2}}34\underline{\underline{5}}$ + $P\overline{1}2\underline{\underline{34}}5$ + $P\overline{1}2\underline{\underline{3}}4\underline{\underline{5}}$ + $P1\overline{23}\underline{\underline{45}}$)
= 0.00011586 < 0.00012

(11) P(4) (= probability that any more-than-triple RM occurs in a word pair)

P(4) = P(5) + Probability of any quadruple matching
= P(5) + ($P\underline{\underline{1}}2345$ + $P1\underline{\underline{2}}345$ + $P12\underline{\underline{3}}45$ + $P123\underline{\underline{4}}5$ + $P1234\underline{\underline{5}}$)
= 0.00000045 < 0.0000005

(12) P(5) (= probability that we have any quintuple RM in a word pair)

P(5) = P12345 = P1 x P2 x P3 x P4 x P5
= 0.0000000004880911565 < 0.0000000005 (= 5^{-10})

The first 100-word list of English and German shows 55 MMs: 3 word pairs show quadruple RMs, 17 word-pairs show triple RMs, and 35 word pairs show double RMs (cf. Ringe 1992: 35-7). The binomial distribution for each type of MMs with P(2), P(3), and P(4), respectively, can be computed by using the formula in (3), as in (13):

(13) The binomial distribution for each type of MMs with P(2), P(3), and P(4)

(a) Distribution for a MM of P(2) [0.01]

no. matches	%	cumulative
0	36.60323413	36.60323413
1	36.97296377	73.5761979
2	18.48648188	92.06267978
3	6.099916581	98.162596361
4	1.494171486	99.656767847
5	0.2897787124	99.9465465594
6	0.04634508026	99.99289163966
7	0.006286345663	99.999177985323
8.	0.0007381693771	99.9999161547001
9.	0.0000762195092	99.9999923742093
10.	0.000007006035694	99.999999380244994
11.	0.0000005790112144	99.9999999592562084
12.	0.00000004337710276	99.99999999263331116
:	:	:

(b) Distribution for a MM of P(3) [0.00012]

no. matches	%	cumulative
0	98.80710014	98.80710014
1	1.185827501	99.992927641
2	0.007044660715	99.999972301715
3	0.00002761838421	99.99999992009921
4	0.00000008037914355	99.9999999.......
:	:	:

(c) Distribution for a MM of P(4) [0.0000005]

no. matches	%	cumulative
0	99.99500012	99.99500012
1	0.004999752506	99.999999872506
2	0.0000001237439364	99.9999999962499364
3	0.000000000002021152	99.999999996251957552
:	:	:

As we can see from the given binomial distributions, the probability of the 55 MMs occurring with the probability P(2) [0.01] is extremely small so it can exclude almost any possibility of chance. This then is precisely the probabilistic evidence for the close non-chance relationship between English and German which can replace the 'strong impression' about the closeness of the language relationship. Furthermore, this evidence is much more decisive than the probabilistic evidence (based on the number of RMs) which Ringe (1992) attempts in vain to provide.

3.2. ENGLISH-LATIN

The numbers of the 'similar' sounds found in the first 100 words of the Swadesh list for English and Latin are given in (14)-(17).[6] The corresponding probabilities P1, P2, P3, P4, and P5 can be calculated in the way described in (2a), as follows:

(14) First position

- 7 pairs of similar sounds in 31 word-pairs
- **P1** = (8/100 x 22/100) + (9/100 x 14/100) + (14/100 x 9/100) + (8/100 x 8/100) + (8/100 x 7/100) + (4/100 x 3/100) + (5/100 x 2/100)
 = 570/10000 = 0.057

(15) Second position

- 1 pair of similar sounds in 2 word-pairs
- **P2** = 3/100 x 2/100 = 6/10000 = 0.0006

(16) Third position

- 2 pairs of similar sounds in 12 word-pairs
- **P3** = (10/100 x 16/100) + (13/100 x 10/100) = 290/10000 = 0.029

(17) Fourth and other positions

- No matches
- **P4** = 0; **P5** = 0

On the basis of the probabilities for any RM in each position, we can calculate the probability that any MM occurs in a given word pair for English and Latin. Here, the

[6] For the first 100-word list, see Ringe (1992: 83-85). As for the 'similar' sounds found in each comparable position, on the other hand, refer to Ringe (1992: 41, 14, 44-47).

calculation of P(2) alone is enough to provide the necessary probabilistic interpretation of the putative relationship.

(18) P(2) (= probability that any multiple RM occurs in a word pair)

$$\begin{aligned}\mathbf{P(2)} &= P(3) + \text{Probability of any double RM in any positions}\\ &= 1 - \{P(0) + P''\}\\ &= 1 - \{P\underline{\underline{12345}} + (P1\underline{\underline{2345}} + P\underline{\underline{1}}2\underline{\underline{345}} + P\underline{\underline{12}}3\underline{\underline{45}} + P\underline{\underline{123}}4\underline{\underline{5}} + P\underline{\underline{1234}}5)\}\\ &= 1 - 0.998297384 = 0.001702616 < 0.002\end{aligned}$$

The first 100 words of the Swadesh list for English and Latin show 9 MMs (cf. Ringe 1992: 47). The binomial distribution for a MM with P(2) [0.002] can be computed, as in (19).

(19) The binomial distribution for a MM with P(2) [0.002]

no. matches	%	cumulative
0	81.8566805	81.8566805
1	16.40414438	98.26082488
2	01.627264824	99.888089704
3	0.106527691	99.994617395
4	0.005176947	99.999794342
5	0.000199193147	99.999993535147
6	0.000006320424	99.999999855571
:	:	:

The probabilistic interpretation of 9 MMs in the English and Latin word is that the occurrence of the given number of MMs is extremely difficult to explain by chance. Even though the non-chance probability for those 9 MMs is not so big as the one for the 55 MMs from English and German, it is still big enough to exclude almost any possibility of chance. Such a probability can be said to reflect our strong impression about the closeness of the non-chance relationship between English and Latin as well as the difference we feel between the relationships of the two pairs of languages (i.e. English and German, on the one hand, and English and Latin, on the other).

3.3. ENGLISH-TURKISH

The first 100-word list of English and Turkish (cf. Ringe 1992: 86-89) shows several 'similar' sounds in some of the comparable positions (i.e. the third and fourth positions). They are given in (20) through (23), along with the corresponding probabilities P1, P2, P3, P4, and P5.

(20) First position (cf. Ringe 1992: 14, 48-9)

- 2 pairs of similar sounds in 8 word-pairs: 6 [b-k] (10-17); 2 [y-s] (2-6)
- **P1** = (10/100 x 17/100) + (2/100 x 6/100) = 182/10000

(21) Second position

- No similar sounds [7]
- **P2** = 0

[7] No RMs are found since there are no initial clusters in Turkish.

The sounds in the third position and their frequencies are given in (22a), the expected chance matchings for all possible pairs are given in table 1 in (22b). and the numbers of matchings actually found are given in table 2 in (22c), as follows:

(22) Third position

(a) Frequencies of consonants

English		Turkish	
Ø	18	r	15
n	15	l	13
t	13	Ø, n, m	9 each
r	10	ğ	8
l	8	t	7
d	6	s, z	6 each
m	5	k	4
k	4	č, p	3 each
θ, s, s, g, ŋ	3 each	ǰ, d	2 each
v	2	b, v, s, h	1 each
p, f, ð, z	1 each		
total	100	total	100

(b) Expected chance matchings in the third consonants, English-Turkish

Tk / Eg	r (15)	l (13)	Ø, n, m (9 each)	ğ (8)	t (7)	s, z (6 each)	k (4)	č, p (3 each)	ǰ, d (2 each)	b, v, s, h (1 each)
Ø (18)	2.7	2.34	1.62	1.44	1.26	1.08	0.72	0.54	0.36	0.18
n (15)	2.25	1.95	1.35	1.2	1.05	0.9	0.6	0.45	0.3	0.15
t (13)	1.95	1.69	1.17	1.04	0.91	0.78	0.52	0.39	0.26	0.13
r (10)	1.5	1.3	0.9	0.8	0.7	0.6	0.4	0.3	0.2	0.1
l (8)	1.2	1.04	0.72	0.64	0.56	0.48	0.32	0.24	0.14	0.08
d (6)	0.9	0.78	0.54	0.48	0.42	0.36	0.24	0.18	0.12	0.06
m (5)	0.75	0.65	0.45	0.4	0.35	0.3	0.2	0.15	0.1	0.05
k (4)	0.6	0.52	0.36	0.32	0.28	0.24	0.16	0.12	0.08	0.04
θ (3)	0.45	0.39	0.27	0.24	0.21	0.18	0.12	0.09	0.06	0.03
s (3)	0.45	0.39	0.27	0.24	0.21	0.18	0.12	0.09	0.06	0.03
s (3)	0.45	0.39	0.27	0.24	0.21	0.18	0.12	0.09	0.06	0.03
g (3)	0.45	0.39	0.27	0.24	0.21	0.18	0.12	0.09	0.06	0.03
ŋ (3)	0.45	0.39	0.27	0.24	0.21	0.18	0.12	0.09	0.06	0.03
v (2)	0.3	0.26	0.18	0.16	0.14	0.12	0.08	0.06	0.04	0.02
p (1)	0.15	0.13	0.09	0.08	0.07	0.06	0.04	0.03	0.02	0.01
f (1)	0.15	0.13	0.09	0.08	0.07	0.06	0.04	0.03	0.02	0.01
ð (1)	0.15	0.13	0.09	0.08	0.07	0.06	0.04	0.03	0.02	0.01
z (1)	0.15	0.13	0.09	0.08	0.07	0.06	0.04	0.03	0.02	0.01

Table 1

(c) Numbers found for matchings of the third consonants, English-Turkish [8]

Tk / Eg	r (15)	l (13)	Ø (9)	n (9)	m (9)	ğ (8)	t (7)	s (6)	z (6)	k (4)	č (3)	p (3)	ǰ (2)	d (2)	b, v, s, h (1 each)
Ø (18)	3	2		3	2	1	2		3	1	1				
n (15)	4	1	1	1	2	2		2		1					1 (v)
t (13)			3	1	1	2	1			2			[2		1 (s)
r (10)	1	2		2				2			1	1			1 (b)
l (8)	3	2				1	1				1				
d (6)		1	1	1				1	1						1 (h)
m (5)		1					1		1					[2	
k (4)	2		1		1										
θ (3)		1				1		1							
s (3)				1	1		1								
s (3)		[2]					1								
g (3)			1		1							1			
ŋ (3)		1			1				1						
v (2)	1					1									
p (1)			1												
f (1)												1			
ð (1)			1												
z (1)	1														

Table 2

(d) Similar sounds based on RMs

similar sounds	(frequencies)	number of matches
[1] t - ǰ	(13 - 2)	2
[2] m - d	(5 - 2)	2
[3] s - l	(3 - 13)	2

Total: 3 pairs of similar sounds in 6 word-pairs

• **P3** = (13/100 x 2/100) + (5/100 x 2/100) + (3/100 x 13/100)= 75/10000

As for the sounds in the fourth position, their frequencies, the expected chance matchings for all possible pairs, and the numbers of matchings actually found are given in (23), as follows:

(23) Fourth and other positions > no RM > no similar sounds

• **P4** = 0; **P5** = 0

(a) Frequencies of consonants

English		Turkish	
Ø	86	Ø	79
d	6	m	12
t	3	n	3
n, k	2 each	d, r	2 each
θ	1	z, k	1 each
total	100	total	100

[8] The numbers in brackets are the matchings (i.e. RMs) which cross the 99th percentile threshold.

(b) Expected chance matchings for the fourth consonants, English-Turkish

Tk / Eg	Ø (79)	m (12)	n (3)	d, r (2 each)	z, k (1 each)
Ø (86)	67.94	10.32	2.58	1.72	0.86
d (6)	4.74	0.72	0.18	0.12	0.06
t (3)	2.37	0.36	0.09	0.06	0.03
n (2)	1.4	0.24	0.06	0.04	0.02
k (2)	1.4	0.24	0.06	0.04	0.02
θ (1)	0.79	0.12	0.03	0.02	0.01

Table 3

(c) Numbers found for matchings of the fourth consonants, English-Turkish

Tk / Eg	Ø (79)	m (12)	n (3)	d (2)	r (2)	z (1)	k (1)
Ø (86)	69	9	3	2	1	1	1
d (6)	5	1					
t (3)	3						
n (2)	1						
k (2)	1						
θ (1)					1		

Table 4

On the basis of the probabilities for any real matches (i.e. P1, P2 and etc.) in the first 100 words of the Swadesh list for English and Turkish, we can calculate the probability that any MM occurs in a given word pair, as in (24) below:

(24) P(2) (= probability that any multiple RM occurs in a word pair)

P(2) = P(3) + Probability of any double RM in any positions
= 1 - {P(0) + P"}
= $1 - \{P\underline{\underline{12345}} + (P1\underline{\underline{2345}} + P\underline{\underline{1}}2\underline{\underline{345}} + P\underline{\underline{12}}3\underline{\underline{45}} + P\underline{\underline{123}}4\underline{\underline{5}} + P\underline{\underline{1234}}5)\}$
= 0.0001365 < 0.00014

$P\underline{\underline{12345}} = 0.9744365$; $P1\underline{\underline{2345}} = 0.0180635$; $P\underline{\underline{1}}2\underline{\underline{345}} = 0$ (< P2=0)

$P\underline{\underline{12}}3\underline{\underline{45}} = 0.0073635$; $P\underline{\underline{123}}4\underline{\underline{5}} = 0$ (< P4=0); $P\underline{\underline{1234}}5 = 0$ (< P5=0)

No MM is found in the first 100 words of the Swadesh list for English and Turkish and this can be verified by the fact that every pair of similar sounds found in the list occurs in a different word pair, as in (25).

(25) Word pairs which show similar sounds

(a) First position (2 pairs of similar sounds in 8 word-pairs)

word pair	meaning	matching	word pair	meaning	matching
bark - kabuk	'bark'	(b - k)	blood - kan	'blood'	(b - k)
belly - karin	'belly'	(b - k)	bone - kemik	'bone'	(b - k)
bird - kus	'bird'	(b - k)	yellow - sarı	'yellow'	(y - s)
black - kara	'black'	(b - k)	you - sen	'you'	(y - s)

(b) Second position (3 pairs of similar sounds in 6 word pairs)

word pair	meaning	matching	word pair	meaning	matching
night - geǰe	'night'	(t - ǰ)	fish - balɨk	'fish'	(s - l)
hot - sɨǰak	'hot'	(t - ǰ)	ash - kül	'ash'	(s - l)
human - adam	'human'	(m - d)			
woman - kadɨn	'woman'	(m - d)			

The binomial distribution for a MM with P(2) [0.00014] is given, as in (26):

(26) Binomial distribution for a MM with P(2) [0.00014]

no. matches	%	cumulative
0	98.60965778	98.60965778
1	1.380728511	99.990386291
2	0.009569788351	99.999956079351
3	0.00004377196013	99.99999985131113
:	:	:

What this binomial distribution in (26) means is that the occurrence of one MM is more than 98.6 % non-chance, which means that out of 1000 hundred-word lists (or out of 100000 word pairs) we can expect one MM in 14 lists. This again means that although finding a MM or two in some of the lists is possible and expected, it will still be very difficult. This is compatible with the fact that no MM was found in the given 100-word list.

In addition, the given probabilistic interpretation supports our expectation about the putative relationship between English and Turkish based on the comparative method. Thus, unlike Ringe (1992), we don't have to appeal to any extra-probabilistic arguments such as historical arguments for English and Turkish.

4. CONCLUSION

In this paper, I have proposed an alternative probabilistic method for determining the (non-)chance relationship between putatively related languages. In particular, the probabilistic evidence (especially, P(2) and the corresponding binomial distribution) based on MMs representing the convergence of similarities has been used to provide a better probabilistic prediction and to deal with problems remained unsolved in Ringe (1992).

In short, supporting and debunking claims or hypotheses about language relationships should be one of the main goals of the probabilistic methods. Considering the fact that the probabilistic method is rarely necessary for closely related languages such as English and German, the further demonstration of validity of the current method as a sifting device in more difficult cases is a pressing need.

REFERENCES

BAXTER, W. H. & RAMER, A. M. 1996. Review of On calculating the factor of chance in language comparison by D. A. RINGE, Jr. 1992. Diachronica 13..2: 371-384.
HONY, H. C. 1957. An Turkish-English dictionary. Oxford: Clarendon.

IZ, F. & HONY, H. C. 1952. An English-Turkish dictionary. Oxford: Clarendon.
MEILLET, A. 1925/1967. The comparative method in historical linguistics (trans. G. B. Ford, Jr.). Paris: Librarie Honoré Champion.
PAULOS, J. A. 1988. Innumeracy: mathematical illiteracy and its consequences. New York: Hill and Wang.
RINGE, D. A., JR. 1992. On calculating the factor of chance in language comparison. Transactions of the American Philosophical Society, vol. 82, Part 1.
_______. 1995. Nostratic and the factor of chance. Diachronica 12: 1. 55-74.
_______. 1996. The mathematics of Amerind. Diachronica 13: 1. 135-154.
_______. 1998. Probabilistic evidence for Indo-Uralic. In SALMONS, J. C. & JOSEPH, B. D. Nostratic: shifting the evidence. Amsterdam Studies in the Theory and History of Linguistic Science IV. Current Issues in Linguistics Theory. pp. 153-97. Amsterdam/Philadelphia: John Benjamins Publishing Company.
WELBY, P. & WHITMAN, N. 1999. Ringe revisited: comments on Ringe's probabilistic comparison method. OSU working papers in linguistics 52 (this volume).